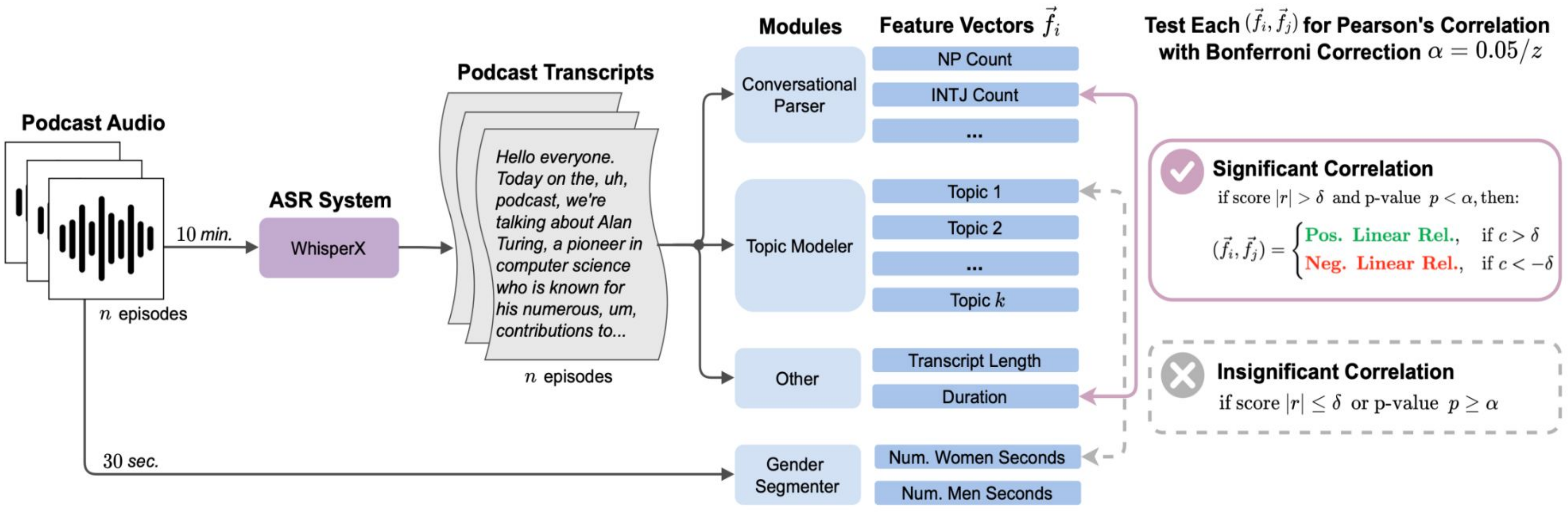
Masculine Defaults via Gendered Discourse in Podcasts and Large Language Models

Maria Teleki, Xiangjue Dong, Haoran Liu, James Caverlee

Women and men's discourse are different.

1 We test for significant correlations between features, including gender and discourse topics.


Podcast Audio
n episodes
10 *min.*
ASR System
WhisperX
Podcast Transcripts
Hello everyone. Today on the, uh, podcast, we're talking about Alan Turing, a pioneer in computer science who is known for his numerous, um, contributions to...
n episodes
30 *sec.*
Modules
Conversational Parser
Topic Modeler
Other
Gender Segmenter
Feature Vectors $\vec{f}_i$
NP Count
INTJ Count
...
Topic 1
Topic 2
...
Topic k
Transcript Length
Duration
Num. Women Seconds
Num. Men Seconds
Test Each $(\vec{f}_i, \vec{f}_j)$ for Pearson's Correlation with Bonferroni Correction $\alpha = 0.05/z$
Significant Correlation
if score $|r| > \delta$ and p-value $p < \alpha$, then:
$(\vec{f}_i, \vec{f}_j) = \begin{cases} \text{Pos. Linear Rel.,} & \text{if } c > \delta \\ \text{Neg. Linear Rel.,} & \text{if } c < -\delta \end{cases}$
Insignificant Correlation
if score $|r| \leq \delta$ or p-value $p \geq \alpha$

2 We find that there are discourse topics that have correlations with women or men.

Topic N	Gender	r	Topic N Word List	Topic N Categories	Topic N Gender
Topic 3	Women Men	0.15 -0.14	women, woman, men, baby, pregnant, girls, men, doctor, health, birth	Content - Pregnancy	Women
Topic 10	Women Men	0.10 -0.12	energy, body, feel, mind, space, yoga, love, beautiful, feeling, meditation	Content - Yoga	Women
Topic 49	Women Men	-0.21 0.17	game, know, think, team, going, mean, play, year, one, good	Content - Sports	Men
Topic 71	Women Men	0.14 -0.14	christmas, sex, girl, hair, love, get, date, girls, let, wear	Content - Dating	Women
Topic 54	Women Men	– 0.12	get, like, know, right, people, going, podcast, make, want, one	Discourse	Men
Topic 60	Women Men	-0.27 0.20	going, know, think, get, got, one, really, good, well, yeah	Discourse	Men
Topic 62	Women Men	0.33 -0.28	like, know, really, going, people, want, think, get, things, life	Discourse	Women

3 This means that, for example, women and men might use a different filler word.

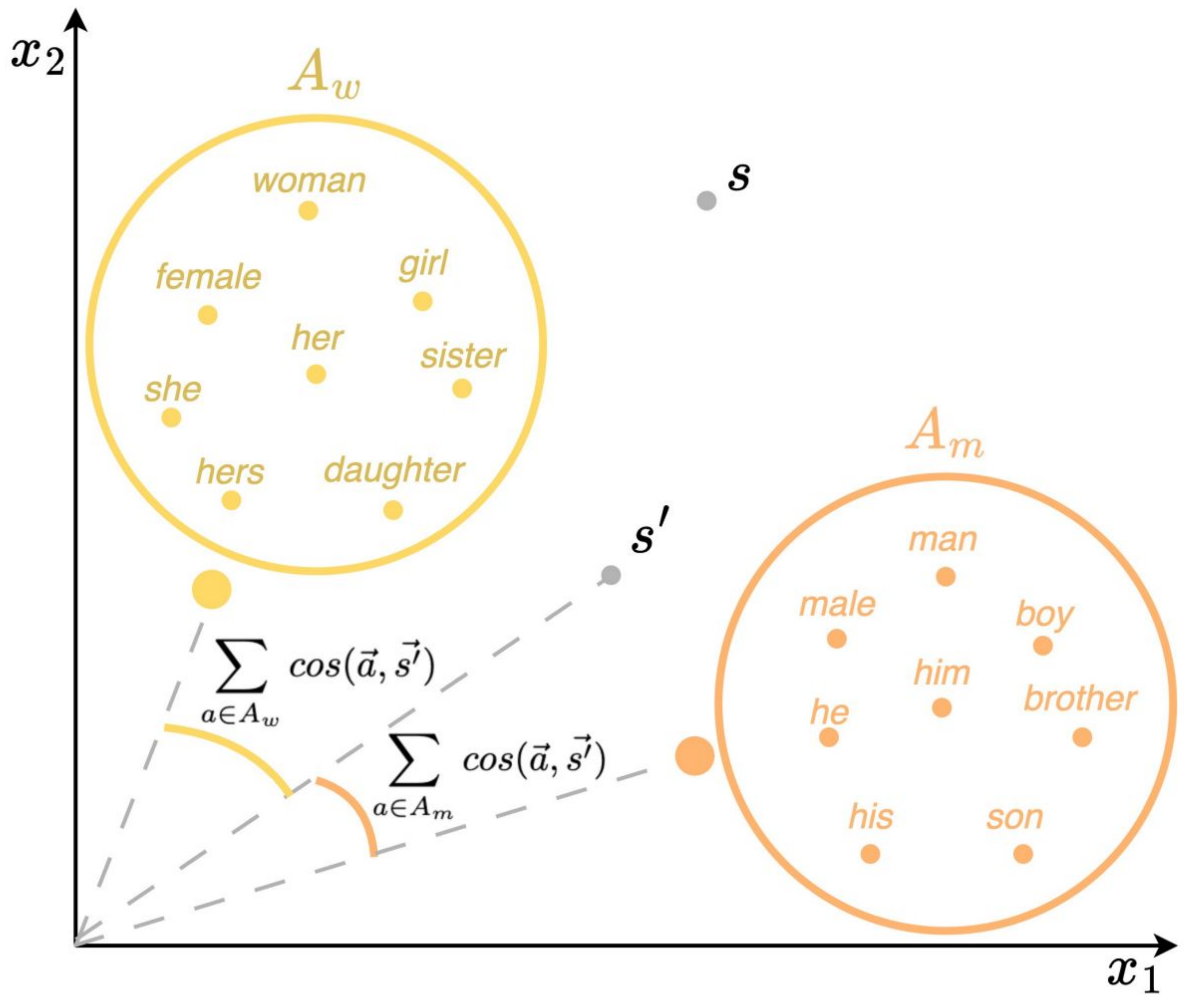
Men: $s =$ *And I was **going**, hey, it's cold outside...*

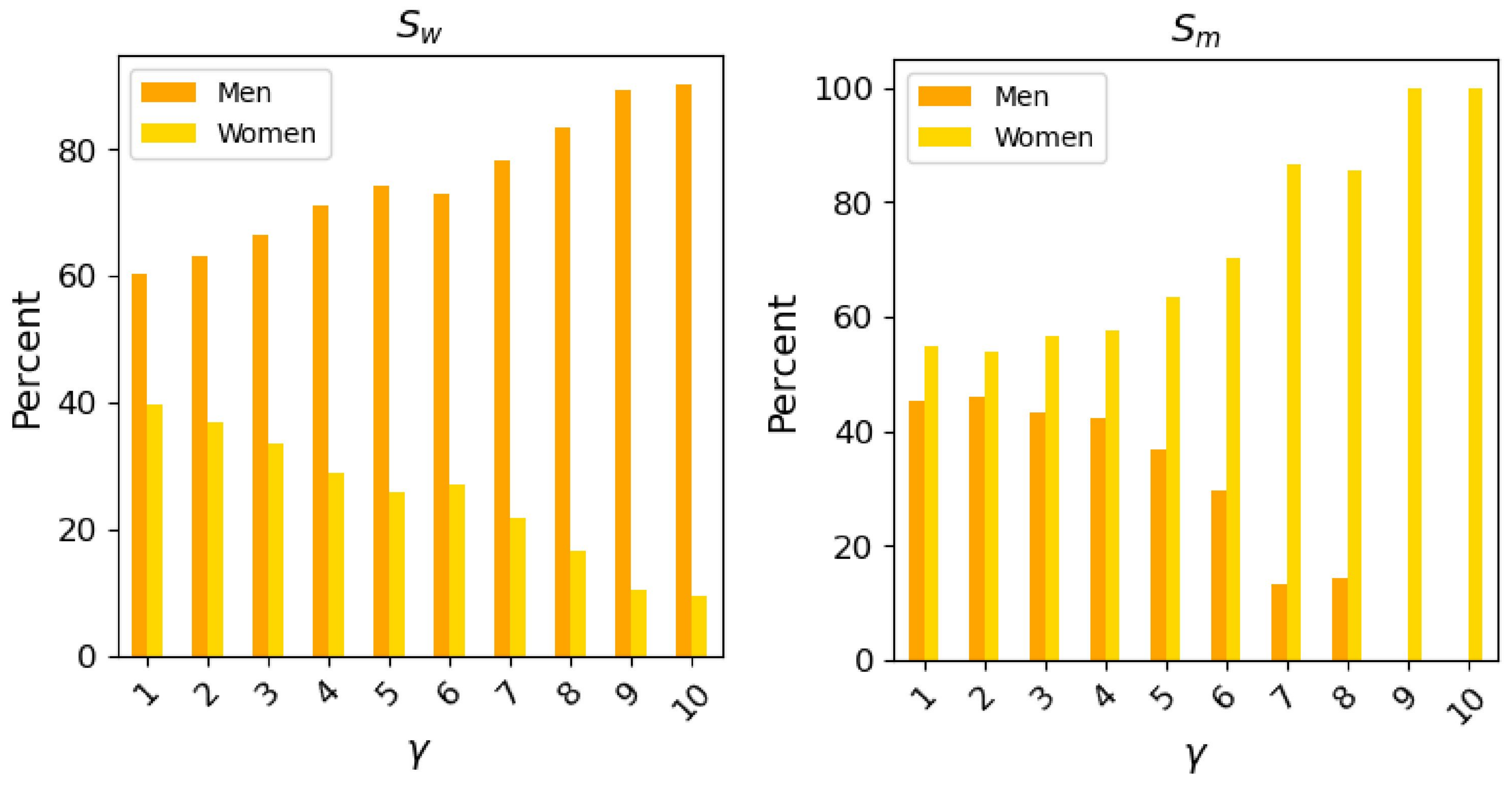
Women: $s' =$ *And I was **like**, hey, it's cold outside...*

These discourse-based masculine defaults are present in LLM embeddings.

4 So we experiment with flipping the gendered discourse words for (s,s') pairs. We measure the movement of s → s' in the embedding space.

5 We find that men have a more stable/robust embedding representation than women w.r.t. discourse words – this is a representational harm & a masculine default.





γ: the # of discourse words flipped in s → s'
Percent: Avg. % of S segments which move closer to A_{m,w} after s → s'

Texas A&M University
Department of Computer Science & Engineering

Paper, code, & more!